Assessment of the psychometric properties of the Graded Care Profile version 2 (GCP2) tool for measuring child neglect.

Robyn Johnson[a,*] and Helen L. Fisher[b]

[a] *National Society for the Prevention of Cruelty to Children, London, UK*

[b] *Social, Genetic & Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK.*

*Corresponding author details:

NSPCC

Weston House

42 Curtain Road

London EC2A 3NH

UK

Tel: +44 020 37729135

E-mail address: robynjohnson@nspcc.org.uk

Acknowledgements

The authors would like to thank:

- The families who agreed for their information to be used for the testing.
- The NSPCC practitioners and managers who took part in the testing and supported the evaluation in a number of ways such as seeking consent from families.
- Priscilla Martens, Executive Director, National Family Preservation Network for permitting the use of NCFAS-G for this testing.
- Dr Robert Bradley at T. Denny Sanford School of Social and Family Dynamics for permitting the use of the HOME Inventory for this testing.

Funding source: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest: none

Abstract

Child neglect is highly prevalent, with extensive effects on children's development, and often long-lasting consequences. It is a complex issue and the identification and assessment of neglect presents particular difficulties for professionals. The Graded Care Profile (GCP) is a tool which can help practitioners to assess neglect, but the updated version of this tool (GCP2) requires psychometric testing. We sought to test the GCP2 with qualified social work practitioners in the UK. Inter-rater reliability was assessed by comparing GCP2 scores between three pairs of practitioners following assessments of 30 children. Concurrent validity was assessed by comparing GCP2 scores to scores obtained using two other validated measures with 15 children. Professionals were surveyed concerning face validity. There was a substantial level of inter-rater agreement for two domains and a moderate level of agreement for the other two domains. There were strong correlations between the GCP2 and the other tools across all domains, and face validity results were positive. In conclusion, although this was a small sample, our findings provide preliminary evidence that the GCP2 has adequate psychometric properties pending further rigorous psychometric testing. It is thus likely to be a useful tool in aiding professionals in the assessment of child neglect.

Key words: Assessment; Child Abuse (neglect); Evidence-Based Practice; Prevention of Child abuse

1. Introduction

It is widely recognised that child neglect is highly prevalent in society. It is the most common form of child maltreatment in several countries (Allnock, 2016), and across the UK, neglect is the most common reason for a child to be the subject of a child protection plan (Jutte et al., 2015). A prevalence study in the UK found that 10% of 11-17-year-olds said they had been severely neglected by parents or guardians (Radford et al., 2011). Neglect has extensive and far-reaching consequences for children's social and emotional development (as discussed in Stevenson, 2007). Children who are neglected can also experience serious and long-lasting consequences, including long-term effects on relationships (Howe, 2011) and an increased risk of mental health problems, including depression and dissociative disorders (Child Welfare Information Gateway, 2009).

However, addressing the problem of neglect is not straight forward. Child neglect is a complex issue, with definitional challenges, and the identification and assessment of neglect presents particular difficulties for professionals. A systematic review by Daniel et al. (2010) found that a range of approaches have been used to define neglect, some focusing on the child's unmet needs while others used narrower definitions. The complexity relates in part to different sub-types of neglect such as physical and emotional; various etiological pathways; confusion and overlap between neglect and other forms of maltreatment such as emotional abuse; and variations in the conceptualisation internationally and cross-nationally, partly relating to the significance of cultural factors. It has also been suggested that the professional role of the definer may influence how they define and understand neglect (Dubowitz et al., 1998). In reaching an understanding of neglect, Stevenson (2007) suggests it is helpful to focus on family dynamics, with a framework which encompasses a wide range of emotional and psychological considerations: *"We should not seek to understand the problems of neglectful parents solely in terms of external pressures on them, but rather in terms of individual and family dynamics, with external pressures as (often powerful) contributing factors."* (p.45)

A survey across English Local Safeguarding Children's Boards (LSCBs) showed inconsistency in the criteria for establishing the presence of neglect; confusion about definitions; fear of making negative value judgments; assumptions about cultural factors and children's resilience; and many examples of children left in damaging situations without support (Gardner, 2008). Learning from serious case reviews (and equivalent reviews in other nations) on neglected children have revealed a failure to identify children's unmet needs, partly due to difficulties professionals face in making sound professional judgements about neglect (Brandon et al., 2010).

Moreover, in 2014 Ofsted carried out a thematic inspection of professional responses to neglect across 11 local authorities. They found neglect practice to be "too variable", with half of assessments being inadequate in some way. However, the use of a standardised approach or tool such as the Graded Care Profile (GCP) enabled social workers *"to apply structure and systematic analysis to very complex situations and to identify key areas of risk"* (Ofsted,

2014, p 17). They went on to say that local authorities that used a standardised approach "*…were more likely to achieve consistency in standards of practice especially if social workers and managers were trained in using the model and managers were effective in quality assuring the standard of work."* (Ofsted, 2014, p 17). In 2012, Barlow et al. carried out a review of structured professional judgment, which found the GCP to be one of only two "promising" assessment approaches, as judged by the evidence of their testing and development. Nonetheless, she pointed to the need for further testing of the GCP, in particular of its validity and reliability.

The GCP (Srivastava & Polnay, 1997) was developed as a practical tool to give an objective measure of the care of children across all areas of need where there were concerns about neglect. The authors state that: "*The GCP2 gives a picture of the quality of care from grade 1 (excellent) to 5 (poor) in all areas of the child's needs. This allows an understanding of how these needs are being met given the family's steady state (normal circumstances), and identifies the areas in which the care is deficient and to what scale*" (p6 Srivastava et al., 2015). It is designed to be used by social workers, health visitors or other professionals involved in evaluating the care of a child. The professional decides which child in the family the tool is used with, and whether scoring is done separately or together with any other children in the family. The professional also decides whether the parental care differentiates enough to warrant scoring each parent separately. Professionals gather information through observations of family life and the home environment, conversations with parents and where appropriate children, as well as health records or other professionals as relevant. The completed tool can be used to inform the professional's decision making and provide evidence of neglect (prevention and/or timely referral where there are fewer concerns or prompt action where there are more serious concerns). The tool is often used again after a period of time to monitor change.

The tool draws on the concept of a continuum of care rather than compartmentalising care into neglectful and non-neglectful. There are five grades, which are based on three different factors: the level of care, commitment to care, and the quality of care. The grades are applied to Maslow's model of human needs (Maslow, 1954): physical, safety, love and belongingness, and esteem (Srivastava & Polnay 1997). There are four domains (areas) in the tool. These were originally: A) Physical care, B) Safety, C) Esteem, and D) Love. Following amendment of the tool in 2015 (see below), the domains were renamed and are now known as: A) Physical Care, B) Care of Safety, C) Emotional Care, and D) Developmental Care.

The originators of the tool carried out inter-rater reliability testing soon after it was developed in 1997. Nursery children were scored independently by two different raters. They found an almost perfect level of agreement in the areas of physical care, safety and esteem, and a substantial level of agreement in love, and concluded that the tool was "*user friendly and gave a measure of care with high inter-rater agreement*" (p.340, Srivastava & Polnay, 1997). However, the testing precluded a range of age groups, as only nursery-aged children were involved. The testing was also limited to health visitors and nursery workers, excluding other agencies, most notably Children's Social Services. In addition, validity testing was not

carried out. There were, therefore, gaps in the original psychometric testing of the measure, and further testing was required.

There have been some positive findings in relation to the GCP, including from the NSPCC (National Society for the Prevention of Cruelty to Children)'s evaluation of the tool (see Johnson & Cotmore, 2015). Practitioners and managers generally reported that it was a useful tool in assessing the child's needs, and explained that it often played an important part in informing their case planning, which could result in their concerns being alleviated or escalated. Professionals were also able to identify several features of the GCP which they felt helped them to identify risks and potential harm more effectively. For example, some said that the scoring process helped them to specify the nature of the neglect and to quantify it. It could also serve to make the neglect more "visible"- to them, but also to parents and others.

However, some limitations of the tool were also identified, and improvements suggested, largely in relation to the language used in the tool. This resulted in the second version of the tool, which is now known as 'GCP2'. Amendments focused on the language used in the tool, gaps in the assessment of neglect, confirmation of scaling and scoring processes, and formatting issues - although the core principles remained the same. This paper presents the findings from the NSPCC's evaluation of the reliability and validity of the updated version of this tool for assessing neglect using qualified social work practitioners.

2. Methods

2.1 Sample

All cases were recruited through NSPCC service centres in England, which received referrals from their respective Local Authorities. All NSPCC service centres that had been previously trained in GCP were invited to take part in the testing. Five service centres expressed interest, and four were selected based on their potential to meet compliance criteria (e.g. sufficient referral rates, practitioner availability). One later withdrew from the testing so three service centres took part. Each service centre manager selected two practitioners (i.e. those who had prior experience of using the GCP2 with at least two families, and those who were able to commit to the amount of time and duration required for the testing). Therefore, three pairs of qualified social work practitioners took part in the testing. Each pair was based at a different service centre. Four pairs of practitioners was deemed an ideal number but this was not possible in practice so three pairs of practitioners took part. However, we felt this was adequate to ensure some generalisability of the findings across practitioners and services.

Practitioners accepted all referrals where there was a concern about neglect. Practitioners sought consent from parents for their scores to be used for the testing, after receiving training and guidance. Testing took place between September 2014 and August 2015.

A case was defined as each GCP2 carried out. Therefore, there could be more than one "case" per family, if the GCP2 was carried out with more than one child in a family. Practitioners were advised in their training to complete GCP2s for each child where there was a concern in relation to neglect. In addition, GCP2s could be completed with more than one parent in a family. Practitioners were asked to submit all GCP2s carried out. A maximum of two GCP2s per family were accepted for analysis, to enhance robust testing through the use of as many independent cases as possible. Only one team submitted more than two GCP2s per family. For this site, cases were selected based on: (i) which child had been selected for validity testing; (ii) aiming to have one GCP2 from each parent of a different child; and (iii) obtaining a representative mix of child age groups.

Scores were checked to ensure that differences in different family members were large enough to be considered as reasonably distinct cases, i.e. that the care of two siblings or by two parents was not identical. The threshold for inclusion was at least a one-point difference across at least two items by at least one rater.

There was an average of six visits to each family by practitioners (ranging from three to 11), including a few unannounced visits in some cases. Evidence was therefore collected over a number of visits across several weeks. All visits were attended by both practitioners in two sites; and 90% of visits in the third site (so 96% across all three sites).

2.2 Ethics

This evaluation was approved by the NSPCC Research Ethics Committee. It was carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki). Practitioners sought written consent from parents after receiving training in this from the NSPCC evaluation team.

2.3 Reliability assessment

The GCP2's inter-rater reliability was assessed using pairs of practitioners, who collected the evidence together to complete the measure but carried out the scoring separately. Three pairs of practitioners (across three teams) used the GCP2 with 10 cases each, so there was a total of 30 cases with paired ratings. For each site, the GCP2 was carried out with five families, thus 15 families in total. In each family, GCP2 scores were analysed in relation to two children, so there was a total of 30 children in the sample.

Practitioners scored each item in each sub-area (see Table 1 for details of the number of items in each sub-area). Once all scores had been submitted for testing purposes, reflective case discussions took place in which moderated GCP2 scores were agreed, in conjunction with the practitioners' line manager. This was the version held on the case file and used for all work with the family. The original unmoderated GCP2 scores were not held on the case

file and were classed as ‘research data’. The approach used was therefore a balance in attempting to meet the needs of the both the reliability testing and good social work practice.

For approximately 96% of all visits, both practitioners were present. However, there were occasions where it was difficult for them to be together in the family home, and in these instances the practitioners described to each other exactly what they had witnessed (but making sure to avoid interpretation) to ensure each had the same information from which to conduct their independent ratings.

2.4 Validity assessment

Concurrent validity was assessed by comparing the scores from the GCP2 with those obtained from other measures. An exercise was carried out to identify tools that matched the domains assessed in GCP2. Two were chosen based on the fact they had previously been validated, and between them closely approximated the domains assessed in the GCP2. A combination of two tools was used because we could not identify a neglect assessment tool that covered all domains within the GCP2.

The first tool was the North Carolina Family Assessment Scale (General version) (NCFAS-G) (Kirk & Ashcraft, 1998), which examines family functioning. The NCFAS–G was developed by the National Family Preservation Network (NFPN) as an instrument designed for family assessment and outcome measures in family preservation services and child welfare. The tool is a structured mechanism for rating and recording observations and judgements in relation to family functioning. The scale is used to determine how a family is functioning. There are 8 domains - Environment, Parental Capabilities, Family Interactions, Family Safety, Child Well-Being, Social/Community Life, Self-Sufficiency, and Family Health. Each domain comprises several subscales which represent aspects of the overall domain. The whole of NCFAS-G was used, because permission was granted on the basis of the tool being used in its entirety.

The second tool was the Home Observation for the Measurement of the Environment Inventory (known as the Home Inventory) (Caldwell & Bradley, 1984, 2003). The Home Inventory is an observer-rated measure designed to evaluate the quality and quantity of stimulation and support available to a child in the home environment. The focus is on the child in the environment, specifically the child as a recipient of inputs from objects, events, and transactions occurring in connection with the family surroundings. There are four different versions of the tool for four different age groups:

- Infant/toddler (IT-HOME) - 0-2 years
- Early childhood (EC-HOME) – 3-5 years
- Middle childhood (MC-HOME) – 6-10 years
- Early adolescent (EA-HOME) – 11+

Permission was obtained to use the two most relevant sections of this tool. A mapping exercise was carried out, whereby the sub-domains and items of each tool were mapped conceptually onto the GCP2 based on the content of the tools. This determined which items were compared during analysis (see Table 2). One practitioner in each pair was assigned to conduct the validity testing (therefore a total of three practitioners). These practitioners completed NCFAS-G and the Home Inventory, in addition to the GCP2. They used the information they had collected during the home visits for GCP2 in order to complete the other tools. The practitioner was asked to complete the NCFAS-G and Home Inventory in relation to one child from each family (the child there was the most concerns about). Therefore, 15 children from 15 different families (5 per team) formed the sample for the validity testing.

Finally, the face validity of the GCP2 was investigated by obtaining feedback through questionnaires from 9 professionals: the 6 social work practitioners who took part in the testing, two of their managers, and one expert in the field of child neglect. Questions focused on the extent to which they felt the GCP2 fulfils its purpose, as set out in the GCP guidance, for example, the extent to which they thought GCP2 addresses all areas of a child's needs and the extent to which they thought it enables professionals to give an objective measure of a child's care. A number of questions were asked, with response options on a four-point scale (to a large extent, to some extent, not much, not at all), with space for free text responses.

2.5 Analysis

For the inter-rater reliability testing, GCP2 scores submitted by pairs of practitioners were compared using Cohen's Kappa as the data were on an ordinal (categorical) scale rather than a continuous scale. Kappa scores were interpreted using the conventions provided by Landis and Koch (1977). The 'internal consistency' of each GCP2 domain was calculated using Cronbach's alpha to see how well the items that should go together in a scale do actually hang together (that is are measuring the same construct). Interpretation of these coefficients was based on the guidelines described by George and Mallery (2003). For concurrent validity, the GCP2 scores were compared to the corresponding scores from both NCFAS-G and the Home Inventory. Spearman's rho correlation was used to assess the extent to which the relevant total scores and subscale scores were correlated. This was used because it is not affected by differences in scaling between measures. In terms of face validity, the quantitative data from the questionnaire was analysed in Snap Professional, and open-ended comments were analysed thematically.

3. Results

3.1 Reliability testing

Approximately half of the 30 children were girls (N=14, 47%), with children ranging in age from infant to 14 years (mean=7 years, SD=4.03). Only two children were not of White British origin. Forty percent of cases were classified as 'Child in Need' (i.e. concerns regarding neglectful parenting already established, child deemed to be at risk of harm) and 60% as 'Child Protection procedures or plans' (i.e. serious concerns regarding neglectful parenting identified, such that the child is deemed to have suffered harm). There were no children where there were 'potential concerns' (i.e. cases where there were no pre-existing and confirmed concerns about neglect). However, across sub-domains, 32% of scores in all cases were scores of 1 or 2, so there was a range in the level of concerns with some good care in several domains identified.

The inter-rater reliability results for each area are presented in Table 3. Agreement appeared strongest for the areas of Physical Care and Developmental Care, as the levels were 'substantial' for the teams as a whole and between 'moderate' and 'almost perfect' for the teams individually (Landis & Koch, 1977). Overall agreement for both Care of Safety and Emotional Care was moderate, although Team A had only slight agreement for Care of Safety. Considering each team separately, there were three Kappa scores that did not reach statistical significance: Team A's ratings of Safety and Team B's ratings of Physical Care and Emotional Care. However, agreement was statistically significant for all teams collectively across all four domains. Internal consistency was excellent for Developmental Care, good for Physical Care and Emotional Care and acceptable for Care of Safety. In terms of complete agreement between raters for each GCP2 area, Area A had the highest percentage of complete agreements (80.0%), followed by Areas D (76.6%), B (70.0%) and lastly C (66.6%).

3.2 Concurrent validity

Six of the 15 children were girls and the vast majority (14/15) were White British. Only three children were aged 5 or under. The average age was 9 years (SD=3.5). Six out of the 15 children were classed as 'Child in Need' and nine as 'Child Protection' cases.

The concurrent validity results are provided in Table 4. Overall there were strong correlations between the GCP2 and the two other tools across all domains. The correlations between the main GCP2 areas and the main related areas from the other tools were strong and statistically significant and ranged from 0.55 to 0.91.

3.3 Face validity

The results were generally positive with participants indicating that the GCP2 fulfils its purpose either to a large extent or to some extent across the vast majority of its aims. However, some mentioned perceived "gaps" in the tool, such as domestic abuse, but others felt that "if used properly" the GCP2 does indeed address all relevant needs. Nonetheless, there were some doubts with regards to the extent to which the GCP2 enables a child's

development or developmental disorder to be assessed, and the extent to which it enables a professional to understand educational, emotional or behavioural outcomes in a child. There was a sense that these were not necessarily an integral part of the GCP2, for example assessment of development or developmental disorder was seen as a "specialist health assessment" for which other measures would be utilised. All participants agreed that the GCP2 measures what it intends to measure (i.e., neglect) to a large extent (n=9/9). One summarised: *"Yes I think the GCP2 provides clear, evidence based information to assist local authorities to make decisions about a child's care, needs and any plans necessary."* (Practitioner). Participants were also very positive regarding the extent to which the GCP2 is consistent with current knowledge and best practices in child neglect, with one manager explaining *"I think it exceeds the knowledge in general social work practice".*

4. Discussion

4.1 Summary

The results from this study suggest that the GCP2 is a valid and reliable measure of child neglect which has moderate to high inter-rater reliability (with strong internal consistency); a high level of concurrent validity; and a fairly high level of face validity. Agreement was statistically significant for all teams collectively across all domains, but not for each team individually for each domain. The lowest level of agreement was from Team A for ratings of Care of Safety. There are two possible explanations for this. Firstly, this team typically received referrals for large families, often with 6 or 7 children. This meant that it was not always possible for the two practitioners to stay together during visits, resulting in one practitioner witnessing safety issues which were not seen by the co-rater. This emerged in the moderation exercise, after GCP2 scores were submitted. Secondly, Team A was the first team to start using the GCP2 for the testing process. In one family in particular, one rater included the garden in the scoring of section B while the other did not. This resulted in some discrepancy in the scoring, as the garden was very unsafe and "worse than the state of the house". However, once this issue had become apparent, all participating practitioners were advised that the garden should be included in scoring, thereby improving the inter-rater reliability for Area B during the course of the testing.

With regards to concurrent validity, the GCP2 correlated well with existing tools that measure aspects of neglect. There were only a small number of items where the correlation was not statistically significant: four of the sub-areas from Area A did not correlate strongly with some of the related areas from the other tools. These were mostly items related to health, and are likely to be due to the fact that the descriptions/definitions of these items are not an exact conceptual match between tools. For example, part of the criteria for judging the Physical Health item in NCFAS-G relates to the child's physical health status, whereas this is not included in the GCP2. In addition, 'environmental risks' in NCFAS-G refers to the "family's living and neighbourhood circumstances" whereas this was compared to GCP2's housing domain which only refers to the home and not the neighbourhood.

The findings from the current sample also indicated that the GCP2 had a fairly high level of face validity, with practitioners generally agreeing that the GCP2 achieves what it sets out to do. However, the results suggest that it may be beneficial to review the extent to which the GCP2 aims to enable the assessment of child development and developmental disorder, as well as educational, emotional and behavioural outcomes. The evidence also suggests that it would be helpful to consider perceived gaps, particularly how the impact of parental problems on children are reflected in the scoring. These recommendations were taken into account by the test developers following our testing: claims relating to measuring developmental disorder were removed and guidance was added in relation to the consideration of the impact of parental factors:

> *"Parental issues can impact on the quality of care they deliver. These are issues such as parental learning difficulties, domestic abuse, parental substance misuse and parental mental health issues, which aren't measured within the GCP2. The practitioner may become aware of these during home visits. It's vital to note these issues as they will assist in understanding where to focus the work with the family. As always, the scoring should only be based on the quality of care delivered by the parent and then explained in the analysis in the report produced. It's important to view the scores in their totality — this will help to explain the overall standard of care the child is experiencing."* (Srivastava et al., 2015, p.16)

4.2 Limitations

There are several limitations that should be considered when interpreting the findings of this study and these mainly centre on the representativeness of the sample. Firstly, the results are based on only a small number of families and thus should be interpreted with caution as statistical power was limited. Additionally, the generalisability of the results is restricted as only 6 practitioners completed the assessments. Secondly, all of the children included were identified as having 'child in need' or 'child protection' status, hence the sample does not include cases at a much earlier stage of need identification. However, as stated in the results section there was a range of scores, indicating that there was very good care in at least some domains in some of the families, so the sample does include cases with different types and levels of concern. The vast majority of children were White British, and the validity testing only included three children under five. These demographic limitations, in particular the small sample size and unrepresentative white British sample, limit generalisability and require addressing in future psychometric testing of this measure.

It is possible that the NSPCC practitioners carried out more family visits than may typically take place when GCP2 is used by a Local Authority. There is, therefore, a potential limitation in generalising to different contexts. However, the decision was taken to carry out testing within NSPCC services for a number of reasons, in particular because this provided the opportunity to carry out a more robust test of the measure. In order to achieve optimal inter-rater reliability testing, both raters in a pair should have the same level of knowledge about the family so that prior knowledge does not skew the scoring. It would have been more difficult to achieve this within a Local Authority, where a lot of cases would typically have

an assigned key worker who could already be familiar with the family. In comparison, when cases were referred from the Local Authority to the NSPCC, NSPCC practitioners had no prior knowledge of the case, with one exception. Therefore, all scoring was determined by evidence collected by both practitioners together for the testing, resulting in a stronger test of the tool.

In addition, given the limited sample size utilised in this study we were unable to perform confirmatory factor analysis or related techniques to confirm the factor structure of the GCP2. We recommend that future studies with larger sample sizes (ideally >100) conduct such tests on this measure. Due to these limitations, particularly the small numbers involved, further testing of the GCP2 tool would be beneficial.

4.3 Testing in a practice context

This testing was carried out in a service setting, in which families were receiving assessment and intervention, as appropriate. As such, the process involved navigating a tension between practice imperatives and the requirements of reliability and validity testing. These tensions were addressed and agreements reached in ways that were comfortable for the practitioners in their practice but which did not undermine the robustness of the testing process. There were two issues that were particularly notable.

The first related to the fact that practitioners would usually carry out an assessment in relation to all children in a family where there are concerns about neglect. However, as explained above, the testing process required cases which could be considered as reasonably distinct cases. There was therefore a process whereby the requirements of the testing were considered in relation to the practice requirements to ensure all children's needs were assessed, as appropriate. This involved discussions with practitioners, managers, and the team carrying out the testing. It was agreed that GCP2 would be carried out on all children who warranted assessment, and that all GCP2 scores would be submitted, with a maximum of two cases per family being used for the analysis. This also meant that each team needed to recruit five families (there were at least two cases in each family), which was achievable within realistic timescales. Therefore, the position reached was satisfactory to all parties.

There was also a negotiation process in relation to the level of discussion that could take place between two practitioners in a pair, and the timing of this. In order to fulfil the requirements for inter-rater reliability testing, it was important that scoring was not unduly influenced by the other practitioner in the pair, which meant that reflection and discussion in relation to the evidence gathered was to be avoided prior to scoring. Practitioners felt, however that this contravened their good practice guidelines, of shared reflection during the course of their assessment. As with the other issue, there was discussion with all parties, and the requirements of both good social work practice and robust reliability testing were considered. It was agreed that the scoring would be carried out at the earliest opportunity, usually after each visit or when there was enough information to score each domain.

Practitioners were then free to discuss the scoring for that domain, but no changes were made to the scores used for analysis.

Another tension centred on practitioners altering their natural way of working for the purposes of the testing. As described above, there were occasions where it was difficult for the pair of practitioners to remain together during assessment visits due to the high proportion of large families in their local area. In these instances, practitioners adhered to the guidance they had been given, i.e. they described to the other what they had seen and heard, but did not reflect on it, offer opinions or interpret the evidence. Although this worked in practice, they commented this this did not come naturally to them and could at time present a challenge because as social workers, they had been taught to analyse and interpret what they had seen, and not simply report "the facts".

4.4 Comparability to original testing

The level of agreement in the current inter-rater reliability testing was slightly lower than in the original testing (Srivastava & Polnay, 1997). However, it is unlikely that this indicates that GCP2 has lower inter-rater reliability than the original GCP. It is difficult to make direct comparisons between the two sets of testing for a number of reasons, including that the tools were implemented differently. In the original testing, scoring was based on one short visit (average 25 minutes). However, in the current testing the evidence was collected over a number of visits, and from different sources including conversations with parents, feedback from children where appropriate, observations in the home and other professionals. In addition, some unannounced visits took place as required. Therefore, with a greater volume and range of evidence, it would seem likely that there was more scope for differing views and different interpretations of the evidence. Considering this, it is encouraging that the level of agreement is generally relatively high in the current testing. In addition, the demographics of the samples differed greatly, notably in terms of the level of concerns and child age groups. It is possible that this sort of tool is implemented differently according to such demographics, and therefore that this could influence the level of inter-rater reliability.

4.5 Implications

Neglect is a complex issue. Amongst other things, assessing neglect requires an understanding of the parent-child dynamic (Stevenson, 2007). Indeed, Crittenden (1993) argued that the parents' style of processing information directly contributes to their ability to perceive and accurately interpret their children's states, and then respond appropriately; in ways which meet their children's needs. The GCP2 offers a tool which can enhance professionals' understanding of the family dynamic and help guide the decision-making process in judging the quality of this interaction. It can also facilitate the collation and conceptualisation of tangible evidence for neglect (e.g., physical care, safety, nutrition, etc.), as part of the broader neglect assessment process.

5. Conclusion

The GCP2 was found to be reliable and valid, albeit in a small sample. This provides preliminary evidence that it has adequate psychometric properties, pending further rigorous testing. It is likely to be a useful tool in aiding professionals in the assessment of child neglect. The NSPCC is leading a roll out of the tool across the UK, which could provide the opportunity for more extensive testing.

References

Allnock, D. (2016) Child neglect: the research landscape. In R. Gardner & D. Howe (Eds.), *Tackling child neglect: research, policy and evidence-based practice* (pp. 101-129). London: Jessica Kingsley

Barlow, J., Fisher, J. and Jones D. (2012) *Systematic review of models of analysing significant harm.* Department for Education Research Report. DFE-RR199.

Brandon, M., Bailey, S. and Belderson, P. (2010) Building on the learning from serious case reviews: a two-year analysis of child protection database notifications 2007-2009.

Caldwell, B. M., and Bradley, R. H. (2003) *Home Observation for Measurement of the Environment: Administration Manual.* Tempe, AZ: Family & Human Dynamics Research Institute, Arizona State University.

Child Welfare Information Gateway (2009) *Understanding the effects of maltreatment on brain development.* Washington, D.C.: United States Department of Health and Human Services.

Crittenden, P.M. (1993) An information-processing perspective on the behaviour of neglectful parents, *Criminal Justice and Behaviour*, **20**, 27-48.

Daniel B, Taylor J, Scott J. (2010) Recognition of neglect and early response: overview of a systematic review of the literature. *Child and Family Social Work* **15(2),** 248–257.

Dubowitz, H., Klockner, A., Starr, R.H., Jr and Black, M.M (1998) Community and professional definitions of child neglect. *Child Maltreatment*, **3**, 235-243.

Gardner, R. (2008) *Developing an effective response to neglect and emotional harm to children.* University of East Anglia and NSPCC.

George, D., and Mallery, P. (2003) *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Boston: Allyn & Bacon.

Howe, D. (2011) *Attachment across the lifecourse: a brief introduction*. Basingstoke: Palgrave Macmillan.

Johnson, R. and Cotmore, R. (2015) *National evaluation of the Graded Care Profile*. London: NSPCC

Jutte, S., Bentley, H., Tallis, D., Mayes, J., Jetha, N., O'Hagan, O., Brookes, H., & McConnell, N. (2015) *How safe are our children?* London: NSPCC.

Kirk, R. S., and Ashcraft, K. R. (1998*) User's guide for the North Carolina Family Assessment Scale (NCFAS).* Chapel Hill, NC: University of North Carolina.

Landis JR and Koch GG. (1977) The measure of observer agreement for categorical data. *Biometrics*, **33**, 159–74.

Maslow, A.H. (1954) *Motivation and Personality*. Harper and Row; New York.

Ofsted (2014) *In the child's time: professional responses to neglect*, No. 140059

Radford, L. Corral, S. Bradley, C. Fisher, H, Collishaw, S. Bassett, C. & Howat, N. (2011) *Child Abuse and Neglect in the UK Today.* London: NSPCC.

Srivastava, O.P., Hodson D. & Fountain R. (2015) *Graded Care Profile Version 2*, NSPCC

Srivastava, O.P., Polnay, L. (1997) Field trial of graded care profile (GCP) scale: a new measure of care. *Archives of Disease in Childhood,* **76,** 337–340

Stevenson, O. (2007) *Neglected children and their families*. 2nd ed. Oxford: Blackwell Publishing.

Table 1: Number of items in each sub-area of the Graded Care Profile version 2 (GCP2)

Domain/Area	Sub Area	Number of items
Physical care	Nutrition	5
	Housing	3
	Clothing	3
	Hygiene	1
	Health	4
Care of Safety	Safety In Parents Presence	5
	Safety when parent is absent	1
Emotional Care	Carer's responsiveness	3
	Mutual Engagement	2
Developmental Care	Stimulation	Age 0-2: 1 Age 2-5: 4 Age 5+: 3
	Approval	1
	Disapproval	1
	Acceptance	1

Table 2: Conceptual mapping of the GCP2 with the NCFAS-G and HOME Inventory

GCP2 domains	NCFAS-G domains	HOME inventory domains*
A. Area of Physical care	D6. Absence/ presence of neglect of children	
A1. Nutrition	G4. Food and nutrition H4. Children's physical health C6. Family routines/ rituals	
A2. Housing	A4. Environmental risks A5. Habitability of housing	
A3. Clothing	A6. Personal Hygiene	
A4. Hygiene	A6. Personal Hygiene	
A5. Health	H4. Children's physical health H7. Family access to health/ mental health care	
B. Area of Care of Safety	D6. Absence/ presence of neglect of children	A total B total
B1. Safety In Parents Presence	D7. Absence/ presence of access to weapons A3. Environmental risks A4. Habitability of housing B6. Controls access to media/ reading material	
B2. Safety when parent is absent	B1. Supervision of children	
C. **Area of Emotional Care** (Nature of attachment)	D6. Absence/ presence of neglect of children	A total B total
C1. Carer's responsiveness	C1. Bonding with children C2. Communication with children	A total B total
C2. Mutual engagement	C1. Bonding with children C2. Communication with children E3. Children's relationship with parents/ caregivers	A total B total
D. Area of Developmental Care	D6. Absence/ presence of neglect of children	A total B total
D1. Stimulation	A7. Learning environment B1. Supervision of children B3. Provision of developmental/ enrichment opportunities B5. Promotes children's education B7. Parents/caregivers literacy C6. Family routines/ rituals C7. Family recreation and play activities E2. School performance F2. Relationships with child care, schools and extracurricular activities E5. Children's relationship with peers	A total B total
D2. Approval	B2. Disciplinary Practices D4. Absence/ presence of emotional abuse of Children	A total B total

D3. Disapproval	B2. Disciplinary Practices C1. Bonding with children C2. Communication with children D3. Absence/ presence of Physical abuse of Children D4. Absence/ presence of emotional abuse of Children	A total B total
D4. Acceptance	C1. Bonding with children C3. Expectations of children	A total B total

*Please note 'A total' refers to the domain total for Responsivity and 'B total' refers to the domain total for Acceptance (or Emotional Climate), depending on the child age group. GCP2 = Graded Care Profile 2nd version, HOME = Home Observation for the Measurement of the Environment inventory, NCFAS-G = North Carolina Family Assessment Scale (General version).

Table 3 Agreement between raters for different GCP2 areas

GCP2 Area	Team A Kappa (95% CI)	Team B Kappa (95% CI)	Team C Kappa (95% CI)	All Teams Kappa (95% CI)	Internal consistency: Cronbach's Alpha
A: Physical Care	0.50* (0.21, 0.79)	0.40, NS (0.16, 0.96)	90% agreement[a]	0.65** (0.43, 0.87)	0.84
B: Care of Safety	0.14, NS (-0.30, 0.58)	90% agreement[a]	0.44* (-0.13, 1.01)	0.56** (0.36, 0.77)	0.73
C: Emotional Care	0.44* (0.06, 0.81)	0.39, NS (-0.07, 0.83)	0.81** (0.46, 1.16)	0.57** (0.50, 0.89)	0.88
D: Developmental Care	0.47** (0.13, 0.80)	0.67** (0.28, 1.05)	0.82** (0.52, 1.12)	0.69** (0.50, 0.89)	0.93

* Significant at the 0.05 level. **Significant at the 0.01 level. CI = Confidence intervals, GCP2 = Graded Care Profile 2nd version, NS = Non-significant (p>0.05). [a]There was 100% agreement in 90% of cases, so 'almost perfect'. Unfortunately there was not enough variance to calculate the Kappa score for this area.

Table 4 Correlations between the GCP2 and related NCFAS-G and HOME main areas

NCFAS-G/Home main areas	GCP2 main areas			
	Area A	Area B	Area C	Area D
Absence/presence of neglect (NCFAS-G)	- 0.77**	- 0.91**	- 0.55*	- 0.64*
A total (HOME)	-	- 0.55*	- 0.82**	- 0.79**
B Total (HOME)	-	- 0.67**	- 0.76**	- 0.72**

*Spearman correlation coefficient is significant at the 0.05 level, **Spearman correlation coefficient is significant at the 0.01 level. GCP2 = Graded Care Profile 2nd version, NCFAS-G = North Carolina Family Assessment Scale (General version), HOME = Home Observation for the Measurement of the Environment inventory.